TIPTEKNO'19
TIP TEKNOLOJİLERİ KONGRESİ
MEDICAL TECHNOLOGIES CONGRESS
3-5 Ekim 2019
Kuşadası / Aydın

Comparison of IMF Selection Methods in Classification of Multiple Sclerosis EEG Data

Soner Kotan
Biomedical Engineering
Istanbul Uni. Cerrahpasa
Istanbul, Turkey

Jeroen Van Schependom and Guy Nagels
Center for Neuroscience, Vrije Uni. Brussel
National MS Center Melsbroek
Brussels - Melsbroek, Belgium

Aydin Akan
Deptartment of Biomedical Engineering
Izmir Katip Celebi University
Izmir, Turkey

***Abstract*—Empirical mode decomposition (EMD) method is a powerful tool in the analysis of of nonlinear and nonstationary signals. It decomposes signals into a number of amplitude and frequency modulated signals namely intrinsic mode functions (IMFs). However, some of these IMFs represents the original signal better while some of them are useless. IMF selection methods are suggested to determine the IMFs which represents the original signal better than other IMFs. In this study, we analyzed the effect of IMF selection methods in classification performance. We compared power based, correlation based and power spectral density based IMF selection methods in the classification of the electroencephalography (EEG) signals, which are collected from subjects with multiple sclerosis. The EEG signals are classified as the patients are being cognitively impaired or intact. k-nearest neighbors, multilayer perceptron neural networks and random forest classifiers are used in classification. The results show that, effect of IMF selection methods on accuracy is changeable in regard to classifier preference.**

***Keywords*—EMD, IMF selection, EEG, MS classification**

I. Introduction

Empirical Mode Decomposition (EMD) is a fully data driven time-frequency analysis method, proposed by Huang et al. [1]. It decomposes the signals into intrinsic mode functions (IMFs), which are frequency and amplitude modulated basis funcions. The method is an effective tool for processing non-linear and non-stationary signals.Wavelet transform (WT) [2] and independent component analysis (ICA) [3] are other widely used signal processing methods in this field. WT requires different wavelet basis functions while EMD is an adaptive method with no basis function requirement [4].

The non-stationary and non-linear signals can be decomposed into IMFs, and this basis functions can reflect the components of the main signal. However, some of these components are not useful, while some of them are reflects the charasteristics of the original signal. Hence, selecting the effective IMFs has been an important process while working with IMFs. Junsheng et al. proposed an IMF selection method based on energy content of IMFs [5]. The method is based on selecting the IMF, which contains the highest amount of energy. Correlation based IMF selection method is another widely method, proposed by Peng et al. [6].Its selection criterion is that the IMF which has the highest correlation coefficient with the original signal should be chosen. Kotan& Akan proposed a power spectral density (PSD) based selection method [7]. Distances between the PSDs of IMFs and the original signal is the base for the method. In this study, we compared the three IMF selection methods in regard to their effect on classification accuracy. We have classified EEG signals collected from multiple sclerosis patients as cognitively impaired/intact. We have used three different classifier to be able to generilaze the results.

II. Methods

A. Empirical Mode Decomposition and Its Multivariate Extension

Empirical mode decomposition is a fully data-driven tool that decomposes non-linear and non-stationary signals into approximately harmonics [8]. Output signals of the process that have slowly varying amplitudes and frequencies are called Intrinsic mode functions (IMF). Sum of IMFs are equal to the multicomponent signal. Shifting process, which is based on envelope extraction is the main part of EMD. The flowchart of EMD algorithm is shown in Fig.1.

Bivariate [9], trivariate [10] and multivariate empirical mode decomposition [11] are the extensions of EMD which are used to decompose multivariate signals. The difference of these extensions is that they use multi-dimensional envelope extraction [12]. Decomposing multivariate signals by using MEMD is more practical and convenient than decomposing each channel by mono EMD. The number of the IMFs obtained is same for every single channel when MEMD used. 10 IMFs obtained from an EEG signal by using MEMD is shown in figure 2.

B. IMF selection methods

In the energy-based method, the IMFs which contain the high amount of energy should be selected due to they are representative for the main signal. The energy content of a signal is equals to sum of the energy contained in all the IMFs which are produced by the MEMD process. It can be given by

$$E_{x(t)} = \sum_{j=1}^{n} E_{c_j(t)} \tag{1}$$

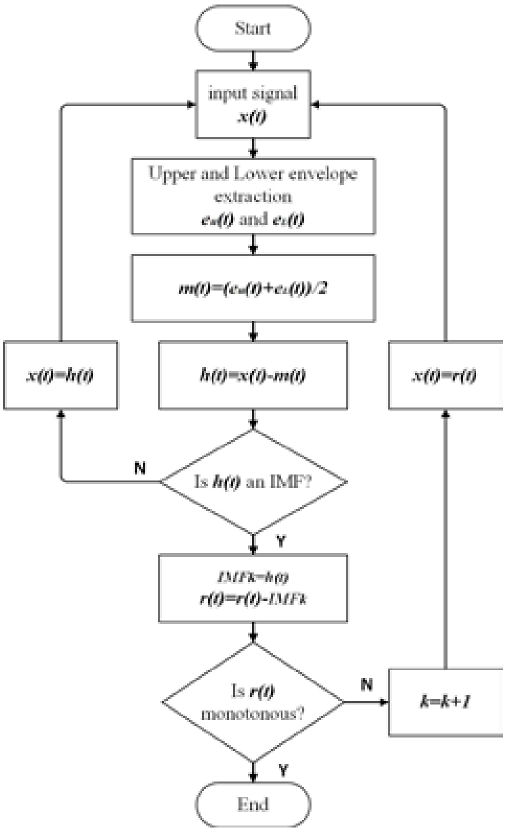


Fig. 1: Flowchart for EMD.

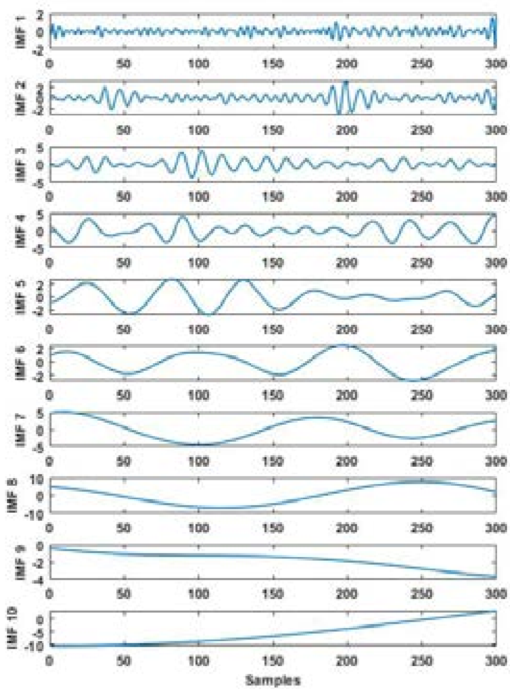


Fig. 2: A sample EEG signal and its IMFs.

where n represents the number of IMFs and $E_{c_j(t)}$ represents the energy content of jth IMF. The energy contained by a sampled IMF is calculated as

$$E_{c_j(t)} = \sum_{i=1}^{N} | \, c_j(i) \, |^2 \tag{2}$$

where N is the number of data samples and $c_j(i)$ denotes the samples of the jth IMF.

Correlation based intrinsic mode function selection method uses the similarity between IMFs and the original signal. Correlation coefficient is a measure of the degree of similarity between two signals, defined as

$$\rho_{x(t)c_j(t)} = \frac{c_{x(t)c_j(t)}}{\sigma_{x(t)}\sigma_{c_j(t)}} \tag{3}$$

where $\sigma_{x(t)}$ and $\sigma_{c_j(t)}$ are the standard deviations of $x(t)$ and jth IMF $c_j(t)$, and $c_{x(t)c_j(t)}$ is the covariance of the signals. The signals have discrete values in practice, thus the correlation coefficient can be given by

$$\rho_{xc_j} = \frac{\sum_{i=1}^{N} \big(x(t) - \bar{x}\big)\big(c_j(i) - \bar{c_j}\big)}{\left[\sum_{i=1}^{N} (x(i) - \bar{x})^2 \sum_{i=1}^{N} (c_j(i) - \bar{c_j})^2\right]^{1/2}} \tag{4}$$

Power spectral density based IMF selection method uses similarity between the PSDs of IMFs and the original signal in selection. Kullback Leibler distance (KL-distance) is employed to calculate the similarity between two signals. The KL-distance between two functions $f(x)$ and $g(y)$

$$D(f \parallel g) = \int f(x) \log \frac{f(x)}{g(x)} dx \tag{5}$$

The similarity of PSDs can be given by

$$PSD_{similarity} = D(PSD(x(t)) \parallel PSD(IMF_i(t))) \tag{6}$$

where PSD is the Furier transfrom of the autocorrelation function of a signal, given in (7). Here in our study, we use Welch method to estimate the power spectral denstiy.

$$S_x(f) = \lim_{t \to \infty} E\left\{ \frac{1}{2T} \left| \int_{-T}^{T} x(t) e^{-j2\pi f t} dt \right|^2 \right\} \tag{7}$$

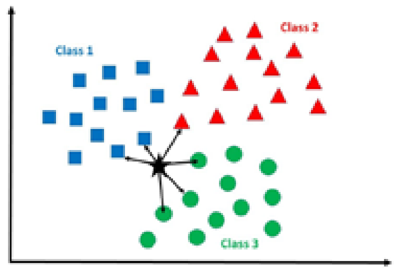
C. Subjects and EEG recordings

The EEG recordings of Multiple sclerosis patients are provided from Melsbroek Edmus database. Digital EEG signals were recorded from 164 MS patients by using Brainlab Measure Station referring to the 10-20 electrode positioning system. The recordings were collected from 21 channels and digitalized at a sampling frequency of 250 Hz. The following electrodes were used for the subsequent analyses: F_{P1}, F_{P2}, F_7, F_3, F_z, F_4, F_8, T_3, C_3, C_Z, C_4, T_4, T_5, P_3, P_Z, P_4, T_6, O_1, O_2, A_1, A_2. EEG signals were filtered in frequency range [1, 30] (Hz) and a 50 Hz notch filter was applied. Mean values of each channel is subtracted from channel values for removal of baseline. The dataset includes EDSS and NSBMS scores, age and MS onset in addition to EEG recordings of patients. Patients are labeled as cognitively impaired/intact by using their test scores.

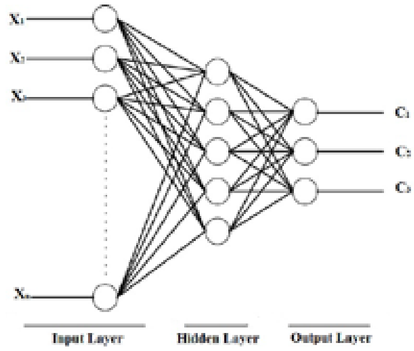
D. Feature extraction and classification

We extracted Hurst exponent, Higuchi fractal dimensions, and kurtosis parameters from the IMFs as features. Hurst exponent is used to measure the self similarity of signals at different scales [13]. Fractal dimensions for a time series is a measure of complexity. Higuchi fractal dimensions, which is one of the various methods for fractal dimension calculation is used [14]. Kurtosis is the fourth-order cumulant to measure the peaked distributions of the random variables.

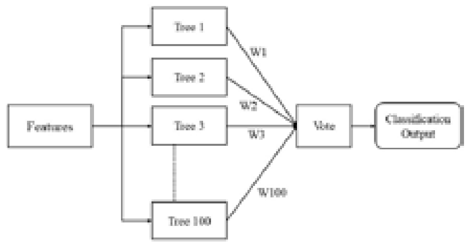
We have employed k-Nearest Neighbor (kNN), multilayer perceptron and random forest as classifiers. kNN algorithm has been regarded as one of the best classification and data mining algorithms due to its simplicity and efficiency [15]. It classifies every new sample in the training set by finding k-nearest neighbors. The classes of the neighbors are used to weigh the class of candidates [16]. Figure 3(a) depits the kNN classifier classify a new sample for k value is 6.
Artifical neural network (ANN) is convenient especially at classification and complex pattern recognition and often used in biomedical modelling, data analysis and diagnose studies [17]. Multilayer perceptron neural networks (MLPNNs), which has hidden layers besides input and output layers is the mostly used model of ANN due to theirsmaller training set requirement [18], ease of implementation [19] and fast operation [20]. The hidden layer weighs input information and transmits it to the output layer.The number of neurons in the hidden layeris not determined by an analytical method. figure 3(b) shows the structure of a MLPNN with single hidden layer.
Random forest (RF) classifier, which consists of many individual classification trees is proposed by Breiman [21]. The classification output of each tree is weighted and used to determine the classification output for random forest. Tree building methods like CART, REPTree, CHAID, etc can be used to build the trees in forest. Average misclassification errors of trees can be used to set the weights of the output of each individual tree. The weights are used in voting procedure todetermine the class. Figure 3(c) shows the structure of a random forest classifier with 100 trees.



(a) kNN classifier for a vector where k value is 6.



(b) The structure of MLPNN with single hidden layer.



(c) The structure of random forest classifier with 100 trees.

Fig. 3: Structures of three classifiers used.

III. RESULTS

21 channel EEG signals, recorded from 50 cognitively impaired and 50 cognitively intact patients are used. Channel C_Z is used as representative channel since it is at the center of the 10-20 EEG placement system. By using MEMD, we have obtained 14 IMFs per channel. Power based, correlation based and PSD based IMF selection methods are applied to select the IMFs that represent the original signal better than other IMFs. Table I shows the IMFs that are suggested to select by three IMF selection methods.

It can be seen from Table I that IMF 4 and IMF 3 are suggested at first and second priority, respectively by three selection methods. We have used first 4 IMFs suggested by three methods in the classifications of EEG signals collected from MS patients, i.e., one class is cognitively impaired, and the other cognitively intact. Table II shows the IMFs which are used in classification process.

Folowing IMF selection process, we extracted Hurst exponent, Higuchi fractal dimensions and kurtosis parameters as features from every selected IMFs. 10-fold cross-validation is utilized for all classification processes.

In our study, we used k-Nearest Neighbour, Multi-layer Perceptron and Random Forest classifiers to detect cognitively impaired and cognitively intact MS patients of EEG signals, by using some nonlinear features namely Hurst exponent, Higuchi fractal dimensions, and kurtosis. Euclidean distance is used as distance measure for kNN (k=1). One hidden layer is used for MLP and the number of nodes are chosen as mean of input and output numbers. The number of trees was 100 for random forest classifier. Every classification process is repeated 100 times and the means and standard deviations are calculated. Classification results are given in Table III.

It can be seen from table that, power based IMF selection method gives the best classification accuracy while k-NN classifier used. While MLP used as classifier, PSD based IMF selection method achieves 70 % accuracy, which is better than other methods. Correlation based IMF selection method gives best classification accuracy with random forest classifier. It can be concluded that, effects of IMF selection methods on accuracy can be changeable in regard to classifier preference.

TABLE I: IMF SUGGESTIONS BY THREE METHODS

IMF Selection	IMF selection methods		
	Power based	Correlation based	PSD based
suggestion 1	4	4	4
suggestion 2	3	3	3
suggestion 3	7	5	5
suggestion 4	9	7	2
suggestion 5	5	8	6
suggestion 6	8	6	7
suggestion 7	6	9	8
suggestion 8	10	2	9
suggestion 9	2	10	10
suggestion 10	11	11	11
suggestion 11	14	12	1
suggestion 12	12	1	12
suggestion 13	13	13	13
suggestion 14	1	14	14

TABLE II: FIRST 4 IMFS SELECTED BY THREE METHODS

	IMF selection methods		
	Power based	Correlation based	PSD based
suggestion 1	4	4	4
suggestion 2	3	3	3
suggestion 3	7	5	5
suggestion 4	9	7	2

TABLE III: CLASSIFICATION PERFORMANCE FOR COGNITIVELY IMPAIRED/OR INTACT TYPE MS PATIENT

	IMF selection methods		
	Power based	Correlation based	PSD based
k-NN	77±2,6	76±2,1	73±2,6
MLP	67±4,2	68±4	70±4,2
RF	73±3,5	75±3,2	74±3,1

IV. CONCLUSIONS

In this study, we have compared the effects of IMF selection methods on classification accuracy. Besides, we used kNN, MLP and Random Forests classifiers to analyze the effects of IMF selection methods on different classifiers. We have used first 4 IMFs suggested by three methods due to use different IMF set from each other. Results show that, effects of IMF selection methods on classification accuracy changes with when different classifiers used and can not be generilazed. The IMF selection methods should be used wiith other classifiers and different kind of signals, to get a better understanding on their effect of classification accuracy.

REFERENCES

[1] N. E. Huang, Z. Shen, S. Long, M. Wu, H. Shih, Q. Zheng, N. Yen, C. Tung & H. Liu, “The empirical mode decomposition and Hilbert spectrum for non-linear and non-stationary time series analysis”, Proceedings of the Royal Society of London, vol. 454, pp. 903-995, 1998
[2] J. Gilles, “Empirical wavelet transform”, IEEE Trans. Signal Process., vol. 61 (16), pp.3999–4010, 2013
[3] A. Hyvärinen, et al., “Independent component analysis: algorithms and applications”, Neural Netw., vol. 45, pp. 411–430, 2000
[4] Ning Bi, et al., “Robust image watermarking based on multiband wavelets and empirical mode decomposition”, IEEE Trans. Image Process., vol. 16 (8) pp. 1956–1966, 2007
[5] C. Junsheng, Y. Dejie, Y. Yu, “Research on the intrinsic mode function (IMF) criterion in EMD method”, Mechanical systems and signal processing, vol. 20 (4), pp. 817-824, 2006
[6] Z. K. Peng, W.T. Peter, F.L. Chu, “A comparison study of improved Hilbert–Huang transform and wavelet transform: Application to fault diagnosis for rolling bearing,” Mechanical systems and signal processing, vol. 19 (5), pp. 974-988, 2005
[7] S. Kotan & A. Akan, “A New Intrinsic Mode Function Selection Method Based on Power Spectral Density”, Medical Technologies National Congress (TIPTEKNO), pp.1-4, 2018
[8] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, H. H. Liu, “The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis”, Proceedings of the Royal Society of London Series A: Mathematical, Physical and Engineering Sciences, vol. 454, pp. 903-995, 1998
[9] G. Rilling, P. Flandrin, P. Goncalves, and J. M. Lilly, “Bivariate Empirical Mode Decomposition”, IEEE Signal Processing Letters, vol. 14, pp. 936-939, 2007
[10] N. ur Rehman, D. P. Mandic, “Empirical Mode Decomposition for Trivariate Signals”, IEEE Transactions on signal processing, vol. 58, pp. 1059-1068, 2010
[11] N. ur Rehman, D. P. Mandic, “Multivariate empirical mode decomposition”, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, vol. 466, pp. 1291-1302, 2009
[12] A. Mert, and A. Akan, “Emotion recognition from EEG signals by using multivariate empirical mode decomposition”, Computer Methods and Programs in Biomedicine, vol. 21, pp. 81-89, 2018
[13] T. Balli, R. Palaniappan, “Classification of biological signals using linear and nonlinear features”, Physiological Measurement, vol.31, pp. 903-920, 2010
[14] T. Higuchi, “Approach to an irregular time series on the basis of the fractal theory”, Physica D: Nonlinear Phenomena, vol.31(2), pp.277-283, 1988
[15] Z. Dong,W. Liang, Y.Wu, M. Pei, and Y. Jia, “Nonnegative correlation coding for image classification”, Science China Information Sciences, vol. 59, pp. 1-14, 2016
[16] G. E. Batista, & M. C. Monard, “An analysis of four missing data treatment methods for supervised learning”, Applied artificial intelligence, vol. 17, pp. 519-533, 2003
[17] M. Sun, and R. J. Sclabassi, “The forward EEG solutions can be computed using artificial neural networks”, IEEE Transactions on Biomedical Engineering, vol. 47, pp. 1044-1050, 2000
[18] E.D. Ubeyli, “Combined neural network model employing wavelet coefficients for EEG signals classification”, Digital Signal Processing, vol. 19, pp. 297-308, 2009
[19] A. Subasi, “Automatic recognition of alertness level from EEG by using neural network and wavelet coefficients”, Expert Systems with Applications, vol. 28, pp. 701-711, 2005
[20] Y. Kocyigit, A. Alkan, H. Erol, “Classification of EEG recordings by using fast independent component analysis and artificial neural network” Journal of Medical Systems, vol. 32, pp. 17-20, 2008
[21] L. Breiman, “Random forests”, Machine Learning, vol. 45, pp. 5-32, 2001